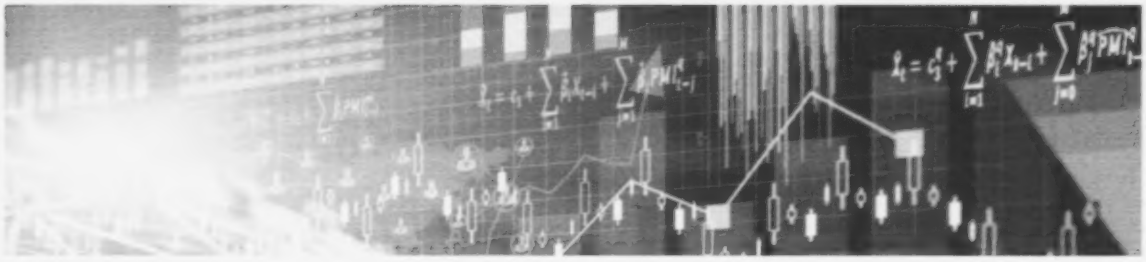
BANK OF CANADA
BANQUE DU CANADA

Staff Working Paper/Document de travail du personnel 2017-43

The Mode is the Message: Using Predata as Exclusion Restrictions to Evaluate Survey Design



by Heng Chen, Geoffrey Dunbar and Q. Rallye Shen

Bank of Canada staff working papers provide a forum for staff to publish work-in-progress research independently from the Bank's Governing Council. This research may support or challenge prevailing policy orthodoxy. Therefore, the views expressed in this paper are solely those of the authors and may differ from official Bank of Canada views. No responsibility for them should be attributed to the Bank.

www.bank-banque-canada.ca

Bank of Canada Staff Working Paper 2017-43

October 2017

The Mode is the Message: Using Predata as Exclusion Restrictions to Evaluate Survey Design

by

Heng Chen,[1] Geoffrey Dunbar[2] and Q. Rallye Shen[3]

[1] Currency Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
hchen@bankofcanada.ca

[2] International Economic Analysis Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
dunbarg@bankofcanada.ca

[3] University of Toronto
rallye.shen@mail.utoronto.ca

ISSN 1701-9397

© 2017 Bank of Canada

Acknowledgements

We thank Kim P. Huynh, Valery Dongmo Jiongo and Tymon Słoczyński for helpful comments. We also thank participants of the Bank of Canada Microeconometrics Workshop, Emory University, Carleton University and Fellowship Exchange for helpful comments and suggestions. All errors are our own.

Abstract

Changes in survey mode (e.g., online, offline) may influence the values of survey responses, and may be particularly problematic when comparing repeated cross-sectional surveys. This paper identifies mode effects by correcting for both unit non-response and sampling selection using the sample profile data (predata) — the individual's number of previous survey invitations, the number of completed past surveys, and the reward points balance. The findings show that there is statistically significant evidence of mode effects in recall and subjective questions, but not for factual ones.

Bank topic: Econometric and statistical methods
JEL code: C8

Résumé

L'utilisation de différents modes d'enquête (questionnaires en ligne et autres) peut influer sur la teneur des réponses et poser notamment problème lors de la comparaison d'enquêtes transversales répétées. Dans cette étude, nous déterminons les effets des modes d'enquête en apportant une solution aux biais de non-réponse et d'échantillonnage. Cette solution passe par l'utilisation du profil de l'échantillon (les « prédonnées »), à savoir le nombre d'invitations reçues précédemment par chaque répondant pour participer à une enquête, le nombre d'enquêtes auxquelles il a déjà répondu et les points de récompense qu'il a accumulés. D'après les résultats, nous disposons de données statistiques probantes concernant les effets qu'ont les différents modes d'enquête sur les questions de rappel et les questions subjectives, mais pas sur les questions d'ordre factuel.

Sujet : Méthodes économétriques et statistiques
Code JEL : C8

Non-Technical Summary

This paper addresses the following question: Do individuals respond differently to the same question when presented with different survey modes (offline vs. online)? We find that for the Bank of Canada's 2013 Methods-of-Payment Survey, there are mode effects for recall questions and for questions eliciting subjective opinions, e.g., whether cash payment method is easy to use. We find no differences for factual questions where the respondent could easily verify the information requested, e.g., the respondent's cash balance. Overall, our empirical results suggest caution when aggregating survey responses for mixed-mode surveys.

In addition, our paper contributes to the statistical methods by applying the predata (information available prior to the specific survey design) to correct for both non-response and sampling bias. Such predata are easy and cheap to obtain from the survey marketing company, and provide enough variation to identify both response choice and sampling strategy, while not confounding the answers of the questionnaire.

1 Introduction

In recent decades, technological advances in telecommunications have given rise to Internet and mobile methods of communication that have been widely adopted. Survey methodologies have evolved to incorporate these recent advances in telecommunications so that surveys are now using novel methods (modes) to both reach and elicit responses. We use data from the Bank of Canada's 2013 Methods-of-Payment (MOP) Survey, which included two survey modes (Chen et al., 2016): a paper-based questionnaire (offline) and an Internet questionnaire (online).[1] We ask: Do individuals respond differently to the same question when presented with different survey modes?

The relationship between survey mode and survey response has been previously studied (see Schonlau et al., 2009); however, this literature does not account, in general, for selection bias due to non-response (Valliant et al., 2013). In this study, we provide an approach to identify the selectivity that generates the non-response bias using sample profile data (hereafter referred to as predata), which we define as information about the past survey experience of individuals in the sample frame. Examples of predata include: the individual's number of previous survey invitations, the number of completed past surveys, and the current reward-points balance earned by completing previous surveys. The predata are unrelated to the primary outcome that interests the researcher, and thus satisfy exclusion restrictions for the participant's response choice.

In addition, this paper also accounts for the quasi-randomized mode assignment by the survey provider to stratify (condition) on both traditional demographic variables and a subset of the survey predata: the individual's number of previous survey invitations and the number of completed past surveys. The individual's number of previous invitations and completions proxy the expected response rates and can be used to stratify the sample to reduce costs (see Groves and Heeringa, 2006). The reward-points balance was not used by the survey company to condition the mode assignment so this provides a separate exclusion that we can disentangle the effects of mode assignment and differential response. The reward-points balance is a suitable

[1] These offline and online questionnaires are identical in language, although there are differences in the presentation of the questions to the survey individual. For example, the online questionnaire is presented sequentially whereas, theoretically, an individual could read and answer offline questions in whatever order they prefer.

candidate because the reward points offered for the survey completion are identical for both modes so there is no marginal advantage for the survey provider to assign an individual to either mode. However, the value of the points to the individual incentivizes survey response.

Given both non-random non-response and mode assignment, we combine Huber (2014) with the availability of the predata to correct for the double selectivity using a nested propensity score. The estimated propensity score to respond is included in the mode assignment propensity score. Since the predata serve as exclusion restrictions, this nested propensity score has sufficient overlap between the two mode treatments. Furthermore, the nested dependence makes our method different from Hamermesh and Donald (2008) who have selection corrections but with two non-nested propensity scores. This is because we allow the mode treatment to affect the response behavior, which is not the case in Hamermesh and Donald (2008).

Our method to identify exclusion restriction differentiates between survey paradata – information collected during the primary data collection process – and what we term predata – information available prior to the specific survey design. See Figure 1 for the relationship among predata, paradata and primary data. In general, the predata exist before a new survey is announced and planned, since the predata consist of convenience sample history up to the time period 0. Once the new survey is announced at time 0, the survey provider can specify the set of paradata variables for collection till time 1, and then both paradata and primary data would be collected in the field before the survey ends at time 2.[2] Survey paradata have been used to correct for survey non-response and measurement errors. Beaumont (2005), Behagel et al. (2014), Kim and Im (2014) and Qin and Follmann (2014) use information on the number of contact attempts to correct for survey non-response. Silva et al. (2016) propose using paradata to correct for measurement errors. One concern with paradata is that they can be costly to collect. For example, repeated attempts to contact survey participants require a larger number of interviewers and can extend the length of time of survey collection. In contrast, survey predata are pre-existing at the time of survey design and do not require additional costs.

In the empirical results, we find that there are mode effects for recall questions and for questions eliciting subjective opinions, e.g., whether a payment method is easy to use. We find

[2]Our definition of *predata* is different from Valliant et al. (2013), in which they refer to the data collected between time 0 and 1.

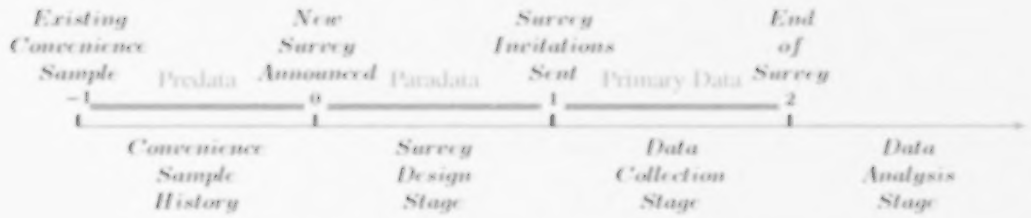


Figure 1: Relationship among predata, paradata and primary data.

no differences for factual questions where the respondent could easily verify the information requested, e.g., the respondent's cash balance. Overall, our empirical results suggest caution when aggregating survey responses for mixed-mode surveys.

While the particular method is specific to our survey mode problem, the basic idea of using the predata as exclusion restrictions provides guidelines to correct for non-response bias more generally. It should be also noted that a related method was employed by Zabel (1998), using characteristics of interviewers and the interview process, and by Hamermesh and Donald (2008), using an affinity identifier – a measure that indicates the degree to which the respondent feels obligated to answer the survey. One methodological contribution of our paper is to show that predata, which are typically readily available for convenience-sample frames, can be used to correct for survey non-response which may, in turn, be exploited to evaluate other survey design effects: e.g., the effect of pre-filling questionnaires based on a respondent's past answers (such as general information on demographics, income or occupation), or the effect of switching from household surveys to individual ones.

Our study is organized as follows. Section 2 discusses how to use the predata as exclusion restrictions for both response behaviors and mode assignment. Section 3 offers the identification strategy for double-selection correction. Section 4 discusses the empirical results and Section 5 concludes.

2 Using predata as exclusion restrictions

This section discusses how predata can be used to account for non-response and the use of predata to account for the specific sampling strategy. We then discuss our data, the 2013 Bank

of Canada Methods-of-Payment (MOP) Survey and its predata variables.

2.1 Predata for non-response

Non-response is a pervasive issue for surveys. The damage appears particularly clearly in randomized trials. Random assignment to treatment creates a treatment group and a control group that are *ex-ante* comparable and representative of the initial population. However, in the presence of non-response, the observed treatment and control groups may no longer be comparable, threatening the internal validity of the experiment. Behaghel et al. (2009) show that non-response bias can be as serious a concern as endogenous program participation by comparing two independent surveys with different response rates. Similarly, there are substantial differences in response rates between the online and offline samples for the MOP survey: a 26% response rate for the paper-based questionnaire versus 1% for the Internet sample.

To adjust for non-response, one possibility is to use exclusion restrictions, which are determinants of the response behavior but not of the outcomes, and that are also obtainable for all invitees in the sampling frame. The practical limitation to relying on exclusion restrictions for the sample selection problem is that credible instruments may not exist. Our contribution is to show that having survey instruments that satisfy the exclusion restrictions to correct for non-response may not be an issue if we have predata for the survey. We acknowledge that predata might be difficult to find in administrative data and that our methods appear more appropriate for convenience frames, such as those normally used by survey firms.

Predata are different from paradata. Paradata usually refer to data about the process by which the survey data are collected. It typically includes the time of day interviews were conducted, how long the interviews took, how many times each interviewee was contacted, the number of attempts to contact the interviewee, and can also include a subjective rating of the reluctance of the interviewee to participate. To correct for non-response using paradata, Beaumont (2005), Kim and Im (2014) and Qin and Follman (2014) use the number of contact attempts to adjust the survey weights, and Behagel et al. (2014) use the number of contact attempts to truncate the treated (or control) group to construct comparable samples. This approach requires that the number of contact attempts be large enough to ensure sufficient

variation across respondents. However, too many attempts may irritate invitees and are often discouraged by the survey provider. It can also be expensive and extend the period of the survey collection. Another paradata variable used for non-response correction is randomized survey rewards (various gifts or monetary incentives) proposed by DiNardo et al. (2006), but they acknowledge that it may be difficult to persuade survey administrators to do so. Shen (2016) examined the use of randomized monetary incentives for the MOP and found negligible predictive power for survey response rates. Both the number of contact attempts and randomized survey-completion rewards tend not to have sufficient variation and often can only identify bounds rather than point estimates.

Predata, by construction, are cheaper to obtain than paradata, both in terms of survey time and cost, and can also be used to randomize survey assignment, since these are the data typically used by survey companies to determine invitation lists. One additional and attractive feature of predata is that the variation in previous survey invitations across individuals is due to quasi-random sampling for those previous surveys. However, the number of completions and reward-points balances are indicators of the willingness to respond and are, therefore, good response predictors. Thus, the predata we exploit appear equally effective as paradata to correct for non-response.

2.2 Predata for sampling strategy

Besides using the predata to correct for non-response *ex-post*, the predata can be used by the survey provider to construct the responsive design *ex-ante* (see Groves and Heeringa, 2006). In many surveys, including the MOP survey, the mode assignment is stratified (conditioned) on both traditional demographic variables and a subset of the survey predata. Conditioning the sampling strategy on the predata is typically cost effective since this data can proxy for the response rates that have been, historically, extremely low for online surveys. Compared with a sampling strategy that does not condition on predata, stratification using predata is expected to result in a more certain sample size and is cheaper for the MOP.

Since the mode assignment is quasi-random after conditioning on the demographics and past survey history, the probability of assignment is varying with respect to both of them.[3]

[3]We control for the sampling selection (stratification) by introducing a set of variables, one for each strata in

Again, the predata used for the sampling strategy also satisfy the exclusion restrictions and have no direct effects on the outcome.

2.3 The MOP Survey

The MOP Survey was conducted by the Bank of Canada to better understand the usage of different payment methods by Canadians and innovation in payment technologies. The topics covered in the questionnaire included consumers' use of cash, credit cards, innovative payment methods like contactless credit and mobile payment, and the respondents' perceptions about the security, convenience and efficiency of these retail-payment methods. In conjunction with an external survey provider, the Bank of Canada targeted a representative national sample for the MOP (see Henry et al., 2015, for a comprehensive overview of the MOP). To minimize survey costs, predata were used to estimate the expected number of responses and to stratify the sample efficiently. Of particular interest for our study, the MOP Survey is a dual-mode survey in which questionnaires were distributed both offline (paper) and online. In this paper, we focus only on a subset of individuals who are included in both the offline and online frames because we cannot correct for frame selection by the survey company.[4]

The data for the MOP study includes a list of socio-demographic variables-such as-age, gender, region, education, income and household size. Table 1 summarizes the predata for the MOP survey.

We partition the predata into two subsets of variables, $Z \equiv (Z^D, Z^R)$, where Z^D are variables related to the assignment to mode and Z, including the additional subset Z^R, are variables used to model the invitee response behavior. Specifically, we define Z^D and Z^R as:

- Z^D: Nbinvites_allstudieson,Nbcompletes_allstudieson, Nbinvites_allstudiesoff and Nbcompletes_allstudiesoff;

which a given set of assignment probabilities is used. Alternatively, one could use design weights; we find that the results are similar.

[4] In the MOP Survey, offline individuals could receive either a 12-page questionnaire or a 16-page questionnaire depending on whether they have participated in an earlier survey on household financial wealth, the Canadian Financial Monitor (CFM). In this paper, we do not differentiate between the 12-page offline and the 16-page offline questionnaires and only define a binary treatment for the mode assignment for two reasons. First, we do not obtain sufficiently detailed predata from the survey company to correct for selection into the CFM frame. Second, the time difference between the 12-page and 16-page survey is reasonably short so we are not concerned that survey fatigue is likely to bias our results.

- Z^R: Nb pointson and Nb pointsoff.

Table 1: The MOP Predata

Variable Name	Description
Nbinvites_allstudieson	Number of *online* surveys to which the respondent has been invited since January 2012.
Nbcompletes_allstudieson	Number of *online* surveys the respondent has completed since January 2012.
Nbinvites_allstudiesoff	Number of *offline (paper)* surveys to which the respondent has been invited since January 2012.
Nbcompletes_allstudiesoff	Number of *offline (paper)* surveys the respondent has completed since January 2012.
Nb pointson	Number of reward points the respondent has earned and not spent since January 2012 through completing online surveys. The number of reward points awarded for each survey is positively related to the length of the survey.
Nb pointsoff	Number of reward points the respondent has earned and not spent since January 2012 through completing paper surveys.

3 Identification based on Huber (2014)

Our goal is to estimate the average treatment effect (ATE) of the survey mode on the observed respondents, where treatment is defined as receiving an online version of the MOP survey, $D = 1$, instead of an offline version, $D = 0$. We label as Y the responses given in the survey.

We are interested in the mode effect D on Y but face the problem that the mode assignment D is selective and that Y is only observed conditional on responses $R = 1$. Hence, we have the partial observation problem as discussed in Poirier (1980): when $R = 0$, we only observed $\{D, R = 0\}$; on the other hand, when $R = 1$, we have observations $\{Y, D, R = 1\}$.

We model the problem using the following three equations:

$$\begin{aligned} R &= I\{h(X, Z^D, Z^R, D) \geq V\}, \\ D &= I\{k(X, Z^D, W) \geq 0\}, \\ Y &= g(D, X, U), \end{aligned}$$

where $\{U, W, V\}$ are unobserved variables with arbitrary correlation between V and U; and $k(\cdot)$ and $h(\cdot)$ are unknown functions. Notice that the response indicator R is assumed to be a separable function between observed $\{X, Z, D\}$ and unobservable scalar V. Recall that the predata variables Z^D for assigning D do not include the reward points Z^R. Both Z^D and Z^R satisfy exclusion restrictions for Y.

To identify the causal effect of D, we follow the potential outcome framework. We denote

$$Y^d = g(d, X, U).$$

The difference $Y^1 - Y^0$ would identify the individual treatment effect, but is unknown because only one treatment status is observed. The additional complication is that the outcomes are observed for a selective subpopulation (only the respondents). We follow Huber (2014) and use a nested propensity score that characterizes both non-random mode assignment and non-random sample selection (respondents being subsample of the entire invite list). The nested propensity score includes the response R propensity score, $P(X, Z^R, Z^D, D) \equiv \Pr\left(R = 1 | X, Z^R, Z^D, D\right)$, along with $\left\{X, Z^D\right\}$ in the treatment D propensity score, which is

$$\Pr\left(D = 1 | X, Z^D, P(X, Z^R, Z^D, D), R = 1\right).$$

This nested dependence of the two propensity scores differentiates our method from Hamermesh and Donald (2008) who consider two non-nested propensity scores. Our method is more general and permits D to affect the response behavior R. For completeness, we list assumptions of Huber (2014) which we use for identification of the ATE:

Assumptions (Huber 2014):

(1a) $Y^1, Y^0 \perp D | \left(X, Z^D\right)$;

(1b) $0 < \Pr\left(D = 1|X, Z^D\right) < 1$.

(2a) $Cov(Z^R, R|X, Z^D, D) \neq 0$ and $Y \perp Z^R|D, X, Z^D$;

(2b) $\Pr(R = 1|D = d) > c > 0$, where $d \in \{1, 0\}$

(2c) $(U, V) \perp (D, Z^R)|\left(X, Z^D, P(X, Z^R, Z^D, D)\right)$;

(2d) $F_V(v)$, the CDF of V, is strictly monotonic in the argument v.

(3) $c < \Pr\left(D = 1|X, Z^D, P(X, Z^R, Z^D, D), R = 1\right) < 1 - c$, where $P(X, Z^R, Z^D, D) \equiv \Pr\left(R = 1|X, Z^R, Z^D, D\right)$.

Assumptions (1a) and (1b) specify the selection-on-observables for D and Y^d. Assumption (2a) allows Z^R in the response model R but not in Y^d. Assumption (2b) rules out that being treated or non-treated perfectly predicts response behavior, while Assumption (2c) imposes that $\left\{D, Z^R\right\}$ are jointly independent of the unobservables $\{U, V\}$, given $\left\{X, Z^D\right\}$ and $P(X, Z^R, Z^D, D)$, with the exogenous variation coming from Z^R. Assumption (2d) is a monotonicity assumption and leads to $\Pr\left(R = 1|X, Z^R, Z^D, D\right) = F_v(h(X, Z^R, Z^D, D))$ under the single-index threshold model R. Finally, the common support of the nested propensity score is required as in Assumption (3).

Let the nested propensity score be

$$\pi\left(X, Z^D, P(X, Z^R, Z^D, D), R = 1\right) \equiv \Pr\left(D = 1|X, Z^D, P(X, Z^R, Z^D, D), R = 1\right).$$

Thus we can condition on $\pi\left(X, Z^D, P(X, Z^R, Z^D, D), R = 1\right)$ instead of $\left\{X, Z^D\right\}$ and $P(X, Z^R, Z^D, D)$. Also, let $\Delta_{R=1} \equiv E(Y_1 - Y_0|R = 1)$ be the ATE in the subpopulation with observed outcomes. We are now able to state Proposition 1 by mapping our model to Huber (2014).

Proposition 1 (Huber, 2014): Under Assumptions (1)-(3) with

$$\begin{aligned} &\pi\left(x, z^D, P(x, z^R, z^D, d), R = 1\right) \\ = \ &\pi\left(x^*, z^{D*}, P(x^*, z^{R*}, z^{D*}, d^*), R = 1\right), \end{aligned}$$

we have

$$\Delta_{R=1} = E\left\{ \begin{array}{c} E\left[Y|D = 1, \pi\left(x, z^D, P(x, z^R, z^D, 1), R = 1\right), R = 1\right] \\ -E\left[Y|D = 0, \pi\left(x^*, z^{D*}, P(x^*, z^{R*}, z^{D*}, 0), R = 1\right), R = 1\right] \end{array} \right\}.$$

To implement Huber (2014), we proceed as follows:

Step 1: estimate $\widehat{P}(X, Z^R, Z^D, D)$ by regressing R on (X, Z^R, Z^D, D);

Step 2: estimate $\widehat{\pi}\left(X, Z^D, \widehat{P}(X, Z^R, Z^D, D), R = 1\right)$ by regressing D on (X, Z^D) and $\widehat{P}(X, Z^R, Z^D, D)$ on the subpopulation $R = 1$;

Step 3: use the sample analogues of Proposition 1.

In our empirical applications, we use parametric specifications based on logit models for Steps 1 and 2, which do not require the choice of smoothing parameters, but can be made arbitrarily flexible by including interactions and higher order terms.

3.1 Robustness check using Huber (2012)

Since the exact selection mechanism underlying R cannot be determined without additional data (see Chen, Felt and Huynh, 2016), as a robustness exercise, we also model R as selection-on-observables.[5] Selection-on-observables does not require exclusion restrictions nor single-index threshold model in the R but is also clearly more restrictive because it assumes that all individual heterogeneity can be captured by the soci-demographic data and the predata. Following Huber (2012), we can estimate $\Delta_{R=1}$ by:[6]

$$E\left\{\begin{array}{c} E\left|Y\right|D=1, \eta\left(\overline{x}, R=1\right), R=1\right| \\ -E\left|Y\right|D=0, \eta\left(\overline{x}^*, R=1\right), R=1\right| \end{array}\right\}$$

where $\eta\left(\overline{X}, R=1\right) \equiv \Pr\left(D=1|\overline{X}, R=1\right)$ with $\overline{X} \equiv \{X, Z^R, Z^D\}$.

Remark: In this study we estimate the mode effect averaged only over the respondents of the MOP and we are unable to generalize this effect to the entire Canadian population. The interpretation of $\Delta \equiv E\left(Y_1 - Y_0\right)$ is not entirely clear in our context. An estimate of Δ would imply that had the entire Canadian population been given surveys (online or offline) and

[5] Another potential direction for the robustness analysis would be to avoid modeling the response process and instead rely on the bound analysis of Lee (2009). Combining predata with bound analysis for double selection is, at present, left for ongoing and future research.

[6] Besides Huber (2012, 2014), Lee and Salanie (2015) and Fricke, Frolich, Huber and Lechner (2015) also discuss how to correct for double selection in different contexts.

all them responded, then their responses for the online sample would differ from those of an offline sample by Δ. In practical terms, this interpretation is unlikely to matter as few surveys obtain 100% response rates. In addition, under most forms of attrition, the identification of Δ requires stronger assumptions than the identification $\Delta_{R=1}$. If these stronger assumptions are not satisfied then Δ is unidentified and identification among respondents appears to be the best feasible alternative, e.g., Newey (2007). Under an additional additive separability condition for the outcome equation ($Y = g(D, X) + U$), we might identify Δ from Das, Newey and Vella (2003) with some restriction(s) on treatment extent, but this separability imposes a conditional constant treatment effect as discussed in Angrist and Fernandez-Val (2010).

The policy relevance of $\Delta_{R=1}$ is precise, if less general, as it refers only to the particular subpopulation of respondents. Nevertheless, we believe that finding significant treatment effects for respondents, $\Delta_{R=1}$, is of interest as it highlights potential pitfalls of mixed-mode survey design.

4 The mode is message

The central goal of this paper is to show that mode effects exist. The exact scale of a mode effect will surely depend on the type of survey conducted.[7] However, we do wish to draw the reader's attention to how potential mode effects may be affected by the nature of the question posed. In the introduction, we described three broad categories of questions: factual/objective, recall and subjective. Factual questions are defined as those that elicit responses of an objective and enumerable nature such as: How much cash do you have on hand? The true answers to these questions can be fact-checked regardless of the survey mode. Next, we define recall questions as those asking the respondent to estimate or remember past choices or actions. The answers to these questions require respondents to contextualize or cogitate and the veracity of an answer is difficult to verify. Finally, we define subjective questions as those that elicit subjective attitudes, such as: How easy or hard is it to use cash? Such a question requires the respondent to define the terms "easy", "hard" and "use" according to her/his particular

[7] As described in Section 3, the MOP is a medium-scale survey that included roughly 50 questions (some with sub-questions). We do not wish to present estimates of the mode effect for all 50 questions, as finding additional evidence of mode effects does not change the message of this paper.

reference. We include one question from each category to demonstrate how mode can interact with responses.

We consider three estimators: Huber (2014), Huber (2012) and naive propensity-score matching only on demographic variables (to highlight the importance of the predata to correct for non-response bias). We compute standard errors by treating the nested propensity score as known. We argue that this approach of computing the standard error would be conservative following from results in Abadie and Imbens (2016).

4.1 Factual/Objective

One focus of the MOP is to estimate the level of cash usage in Canada. The MOP directly asks respondents the level of cash they currently have on hand. The response to this question can be easily verified as respondents can simply look in their wallet or pocket to ascertain the correct answer. We report our estimated mode effect for cash on hand in Table 2. We see no significant difference in the response conditional on mode and this is consistent across the three estimators we consider. These estimated responses are qualitatively (though obviously not quantitatively) similar. In brief, we find no evidence of a mode effect for cash on hand.

Table 2: Mode Effect for Cash on Hand

	Huber (2014)	Huber (2012)	Naive
ATE	-26.7	3.78	-7.91
Std Err	(16.4)	(25.3)	(17.9)
Obs	1043	1043	1043

4.2 Recall

The MOP is also designed to estimate the level of usage of payment methods in Canada. We estimate the mode effect for the number of times cash was used by the respondent in the previous month. The estimated treatment effect for mode is significant for all estimators, although the magnitude of the estimate depends on the particular method used. Our preferred method, following Huber (2014), is roughly one and a half times as large as the naive estimate, which is

based only on demographic propensity-score matching. The estimated effect of mode assignment to an online survey is to reduce the reported number of cash transactions by approximately three times per month in comparison with an offline survey; see Table 3.[8]

Table 3: Mode Effect for Cash Usage

	Huber (2014)	Huber (2012)	Naive
ATE	-3.48**	-3.00***	-2.26**
Std Err	(1.41)	(0.85)	(0.99)
Obs	1040	1040	1040

4.3 Subjective

The final type of question we consider is subjective, by which we mean that the survey question solicits a subjective opinion regarding the respondent's feelings. The MOP asked: [i]n your opinion how easy or hard is it (or would it be) to use cash payment in Canada? We observe that this question requires the respondent to define the terms "ease", "hard" and "use" and that the response to this question should depend on the respondent's particular frame of reference for these terms. For a subjective question, the respondent must evaluate what is meant by "use" and this context is likely to be person-specific. For example, does "use" refer to any or all of:

- Online?
- At a small store?
- At a large retailer?
- At a gas station?
- At a restaurant? ...

... and how does a respondent decide to weigh ease of use across these possibilities? We observe that our estimated mode effect is significant, particularly for our preferred specification,

[8]We do not observe a significant mode effect for other payment methods; however, these questions appear to be affected by item non-response and we are cautious about interpretation.

Huber (2014). In general, online respondents appear to view cash payment as being slightly harder to use than offline respondents do.

Table 4: Mode Effect for Ease of Use

	Huber (2014)	Huber (2012)	Naive
		Cash	
ATE	-0.12**	-0.16***	-0.073
Std Err	(0.047)	(0.051)	(0.056)
Obs	1044	1044	1044

4.4 Discussion

There are two hypotheses about why survey mode may influence responses above and beyond response rate:

1. Survey questions may be framed by a respondent's physical location/times/states, and these locations/time/state may differ across modes.

2. The device used by the mode (e.g., paper or computer) may influence the cognitive process for response.

We find statistically significant evidence of mode effects in recall and subjective responses to identical questions. Finding response differences for recall questions suggests that respondents may require additional prompting or memory aid to ensure no bias in responses. Interestingly, we find some subjective response differences for questions eliciting individual sentiment. We infer that survey mode may influence an individual's cognitive process for emotional responses (supporting the second hypothesis). However, there are no response differences across modes for factual questions, suggesting that such questions are location/time/state independent (ruling out the first hypothesis).

5 Conclusion

In order to infer the impacts of modes on response from these survey results one needs to account for possible non-response bias. To do so, we develop a method of identifying the

selectivity that is inherent in any social-science sampling procedure and use an identifier to adjust the estimates. The identifier we use is the predata. The predata can, we believe, be developed analogously in many surveys, with the results perhaps particularly applicable for inferring the importance of non-response bias.

References

Abadie, A., & Imbens, G. W. (2016). Matching on the estimated propensity score. *Econometrica*, 84(2), 781-807.

Angrist, J., & Fernandez-Val, I. (2010). Extrapolate-ing: External validity and overidentification in the late framework (No. w16566). National Bureau of Economic Research.

Beaumont, J. (2005). On the use of data collection process information for the treatment of unit nonresponse through weight adjustment. *Survey Methodology*, 31(2), 227.

Behaghel, L., Crepon, B., Gurgand, M., & Le Barbanchon, T. (2015). Please Call Again: Correcting Nonresponse Bias in Treatment Effect Models. *Review of Economics and Statistics*, 97(5), 1070-1080.

Chen, H., Henry, C., Huynh, K. P., Shen, Q. R., Vincent, K. (2016). Measuring Consumer Cash Holdings: Lessons from the 2013 Bank of Canada Methods-of-Payment Survey. *Survey Practice*, 9(4).

Chen, H., Felt, M.-H. and Huynh, K. P. (2016), Retail payment innovations and cash usage: accounting for attrition by using refreshment samples. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180(2), 503-530.

Henry, C. S., Huynh, K. P., & Shen, Q. R. (2015). 2013 Methods-of-Payment Survey Results. Bank of Canada Staff Discussion Paper 2015-4.

DiNardo, J., McCrary, J., & Sanbonmatsu, L. (2006). Constructive proposals for dealing with attrition: An empirical example. NBER Working Paper.

Fricke, H., Frölich, M., Huber, M., & Lechner, M. (2015). Endogeneity and Non-Response Bias in Treatment Evaluation: Nonparametric Identification of Causal Effects by Instruments. IZA Discussion Paper No. 9428.

Groves, R. M., & Heeringa, S. G. (2006). Responsive design for household surveys: tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society: Series*

A (Statistics in Society), 169(3), 439-457.

Hamermesh, D. S., & Donald, S. G. (2008). The effect of college curriculum on earnings: An affinity identifier for non-ignorable non-response bias. *Journal of Econometrics*, 144(2), 479-491.

Huber, M. (2012). Identification of average treatment effects in social experiments under alternative forms of attrition. *Journal of Educational and Behavioral Statistics*, 37(3), 443-474.

Huber, M. (2014). Treatment evaluation in the presence of sample selection. *Econometric Reviews*, 33(8), 869-905.

Kim, J. K., & Im, J. (2014). Propensity score adjustment with several follow-ups. *Biometrika*, 101(2), 439-448.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, 76(3), 1071-1102.

Lee, S., & Salanié, B. (2015). Identifying effects of multivalued treatments. CEPR Discussion Paper No. DP10970.

Newey, W. K. (2007). Nonparametric continuous/discrete choice models. *International Economic Review*, 48(4), 1429-1439.

Poirier, D. J. (1980). Partial observability in bivariate probit models. *Journal of Econometrics*, 12(2), 209-217.

Qin, J., & Follmann, D. A. (2014). Semiparametric maximum likelihood inference by using failed contact attempts to adjust for nonignorable nonresponse. *Biometrika*, 101(4), 985-991.

Schonlau, M., Van Soest, A., Kapteyn, A., and Couper, M. (2009). Selection bias in web surveys and the use of propensity scores. *Sociological Methods and Research*, 37(3), 291-318.

Da Silva, D. N., Skinner, C., & Kim, J. K. (2016). Using Binary predata to Correct for Measurement Error in Survey Data Analysis. *Journal of the American Statistical Association*, 111(514), 526-537.

Shen, Q. R. (2016). Recruitment for a Financial Survey: the Impact on Response Rates of Incentives, Demographic Variables and Past Survey Experience. Mimeo.

Valliant, R., Dever, J. A., Kreuter, F. (2013). *Practical tools for designing and weighting survey samples*. New York: Springer.

Zabel, J. E. (1998). An analysis of attrition in the Panel Study of Income Dynamics and the Survey of Income and Program Participation with an application to a model of labor market behavior. *Journal of Human Resources*, 479-506.